Gene-Covariate Interaction between Dysplastic Nevi and the *CDKN2A* Gene in American Melanoma-prone Families[1]

Alisa M. Goldstein,[2] Maria Martinez, Margaret A. Tucker, and Florence Demenais

Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20892-7236 [A. M. G., M. A. T.], and Institut National de la Santé et de la Recherche Médicale, Hôpital St-Louis, Paris, France [M. M., F. D.]

Abstract

The *CDKN2A* gene has been implicated in cutaneous malignant melanoma pathogenesis. Although *CDKN2A* mutations confer substantial risk for melanoma, clinicoepidemiological covariates including dysplastic nevi (DN), total nevi, and solar injury also enhance melanoma risk. To examine the relationship between *CDKN2A* and these three risk factors, we conducted combined segregation/linkage analysis using the class D regressive logistic model, as implemented in the computer program REGRESS. Genetic and covariate data were collected on 20 American melanoma-prone families, 13 of which had cosegregating *CDKN2A* mutations. Two types of analyses were conducted. The missing-indicator method used a missing-value indicator, set to 1 for unknown and 0 for known covariate status, and a second variable set to 1 for exposed and 0 for unexposed or unknown. The second method, complete-cases method, coded subjects with missing covariates as unknown for the affection status. The results for both analyses were very similar. Overall, there was a significant improvement in the likelihood when DN, total nevi or both covariates were added to the base model, which included dominant transmission of the *CDKN2A* gene and a linear increase of risk with the logarithm of age on the logit scale. In contrast, inclusion of solar injury did not significantly improve the likelihood for the base model. Significant evidence for a gene-covariate interaction was detected between DN and *CDKN2A* when DN was the only covariate in the model (missing-indicator method or complete-cases method) or when both DN and total nevi were in the model (complete-cases method only). Interestingly, in both methods, the odds ratio (OR) for DN was greater in subjects without mutations (OR, 20.1; 95% confidence interval, 4.8–92.8) *versus* those with *CDKN2A* mutations (OR, 3.3; 95% confidence interval, 1.1–10.0; complete-cases method). The *CDKN2A*-DN interaction illustrates the complex etiology of melanoma and needs to be confirmed in a larger sample of families.

Introduction

CMM[3] is a potentially fatal form of skin cancer, the etiology of which is heterogeneous and complex. Approximately 10% of malignant melanomas develop in individuals with a familial predisposition (1) and often in association with clinically dysplastic or atypical nevi (2).

To date, two genes have been implicated in melanoma pathogenesis. The first, *CDKN2A*, encodes a low molecular weight protein, p16, that inhibits the activity of the cyclin D1-*CDK4* complex (3). This complex phosphorylates the retinoblastoma protein, allowing the cell to progress through the G_1 cell cycle checkpoint. Thus, p16 acts as a tumor suppressor and negatively regulates cell growth by arresting cells at G_1. Germ-line *CDKN2A* mutations have been detected in 10–25% of melanoma-prone families that have been examined from North America, Europe, and Australia. In contrast, the second melanoma gene *CDK4* acts as an oncogene (4), and germ-line mutations have been detected in only three melanoma-prone families worldwide (5, 6). Other genetic factors remain to be identified.

Although *CDKN2A* and *CDK4* confer substantial risk for melanoma, other environmental and genetic factors are likely involved in the pathogenesis of familial melanoma. Epidemiological studies of melanoma suggest that exposure to sunlight is the major environmental risk factor, although the exposure response relationship appears complex with intermittent sun exposure likely to be more important for risk than total lifetime exposure (reviewed in Ref. 7). The major host factors associated with melanoma are increased numbers of melanocytic nevi, both clinically banal and atypical (dysplastic; Refs. 7 and 8). Previous examination of 13 melanoma-prone families with *CDKN2A* mutations, a subset of the families in the current study, showed results consistent with that seen in epidemiological studies (9), *i.e.*, the same types of clinical and environmental risk factors were observed.

Specifically, examination of the relationship between clinical, environmental, and genetic factors and the risk of melanoma in these 13 families with *CDKN2A* mutations revealed that several variables, in addition to *CDKN2A* mutation status, appeared to enhance disease risk (9). The exposures that most consistently increased melanoma risk were TN, the presence of DN, and evidence of SI, as measured on the backs of shoulders

Received 9/1/99; revised 7/7/00; accepted 7/18/00.
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
[1] This work was partly supported by Fondation de France Grants 3D180B and HDH01E and by the Association pour la Recherche sur le Cancer Grant 1335.
[2] To whom requests for reprints should be addressed, at Genetic Epidemiology Branch/National Cancer Institute, Executive Plaza South, Room 7004, 6120 Executive Boulevard, MSC 7236, Bethesda, MD 20892-7236. Phone: (301) 496-4375; Fax: (301) 402-4489; E-mail: goldstea@exchange.nih.gov.

[3] The abbreviations used are: CMM, cutaneous malignant melanoma; CDK, cyclin-dependent kinase; TN, total nevi; DN, dysplastic nevi; SI, solar injury (or damage); MI, missing-indicator; CC, complete-cases; OR, odds ratio; CI, confidence interval.

Table 1 Number of CMM patients, family members, and distribution of risk factors in melanoma-prone families

Family	Confirmed cases of CMM	Median age at CMM Dx	Total members	*CDKN2A* alteration	Nevi[a]			DN[a]			SI[a]		
					Exp	Unexp	Unk	Exp	Unexp	Unk	Exp	Unexp	Unk
A	8	38	53	23ins24	12	32	9	16	32	5	22	22	9
D	7[b]	34	41	Arg58Ter	9	21	11	13	15	13	15	14	12
E	4[b]	25	26	Asn71Ser	2	22	2	9	14	3	14	10	2
F	8[b]	34	51	Arg87Pro	9	36	6	18	23	10	10	35	6
G	7	31	29	Gly101Trp	5	20	4	11	12	6	8	18	3
H	3	32	9	Gly101Trp	3	5	1	6	2	1	2	6	1
I	5[b]	27	15	Gly101Trp	8	3	4	8	1	6	7	3	5
J	7[b]	39	18	Val126Asp	6	12	0	8	8	2	5	12	1
K	5[b]	36	16	Val126Asp	2	12	2	8	6	2	6	8	2
M	12	38	40	Ala148Thr	7	13	20	13	12	15	9	8	23
N	7[b]	35	21	167del31	3	10	8	5	5	11	7	6	8
P	12	38	68	240del14	0	40	28	7	28	33	18	22	28
Q	3	31	8	IVS2+1	1	5	2	3	2	3	4	2	2
T	4[b]	29	17		0	13	4	7	6	4	5	8	4
U	3	23	9		1	6	2	2	5	2	4	3	2
V	3[b]	42	9		2	7	0	6	2	1	5	4	0
W	3	55	23		5	10	8	10	5	8	15	0	8
X	2	43	13		6	5	2	5	6	2	5	6	2
Y	2	60	6		2	3	1	3	2	1	3	1	2
Z	6[b]	38	35		12	12	11	10	12	13	11	13	11

[a] Exp, exposed; Unexp, unexposed; Unk, unknown.
[b] Includes 1 patient with melanoma-in-situ (MIS).

(9, 10). To search for interactions between the *CDKN2A* gene and these clinicoepidemiological risk factors, we conducted combined segregation/linkage analysis using the regressive models (11–13) in 20 American melanoma-prone families.

Subjects and Methods

Family Data. The study subjects were drawn from 20 families in which there was a history of invasive melanoma in at least two first-degree relatives. The families were referred to the National Cancer Institute through health-care professionals or self-referrals. All of the families were Caucasian and resided in various regions of the United States. The families have been followed prospectively for 4–22 years.

Written informed consent was obtained prior to participation under an Institutional Review Board-approved protocol. All family members willing to participate in the study were clinically evaluated. Clinical evaluation of family members and spouses included complete skin examination, routine medical history, and phlebotomy to obtain lymphocytes. Variables recorded during the clinical examination included the type, distribution, and TN, and evidence for SI (defined as clinical assessment of absence/presence of solar keratosis, thinning of epithelium, idiopathic guttate hypomelanosis, wrinkling, and telangiectasia on backs of shoulders). For subjects who had multiple clinical evaluations, variables for the current study were taken from the most recent examination during which covariate data were recorded. All diagnoses of melanoma were confirmed using histological review of pathological material, pathology reports, or death certificates. Invasive and *in situ* cutaneous melanomas were classified as melanoma. Only confirmed melanomas were included in the analyses.

Table 1 shows the numbers of CMM cases, median age at CMM diagnosis, total family members, and specific alteration by family. The families ranged in size from 6 to 68 members with 2–12 CMM cases. The 20 families included 101 invasive and 10 *in situ* CMM cases and their 396 relatives (275 blood relatives and 121 spouses). Thirteen of the families had germline *CDKN2A* mutations that cosegregated with melanoma (9, 14). The mutations included one nonsense (Arg58Ter), one splice-donor-site (IVS2 + 1), five missense (Val126Asp, Gly101Trp, Arg87Pro, Asn71Ser, and Ala148Thr) mutations, one insertion (23ins24) and two deletions (234del14, 167del31). Melanoma susceptibility gene(s) have not yet been identified in the other seven families; these seven families were negative for both *CDKN2A* and *CDK4* mutations. In addition, these seven families did not show statistically significant evidence for linkage to chromosome 9p21, the location of the *CDKN2A* gene.

Regressive Models. Combined segregation/linkage analysis of CMM was conducted by use of the class D regressive logistic model (11) extended to take into account variable age at diagnosis of disease (15) and linked marker loci (12). The regressive models are constructed by specifying a regression relationship between each person's phenotype (affected/unaffected with CMM) and a set of explanatory variables, including the person's major genotype, the phenotype of older relatives (to take into account family dependence of unspecified origin, genetic, and/or environmental), and measured covariates. The cotransmission of CMM and the *CDKN2A* locus was analyzed as follows. When considering the cosegregation of a trait (Y) and a marker (M), the likelihood of a family can be written as:

$$L(\text{family}) = \sum_{g_Y,\, g_M} P(g_Y, g_M)\, P(Y, M | g_Y, g_M, X)$$

$$= \sum_{g_Y,\, g_M} P(g_Y, g_M)\, P(Y | g_Y, X) \quad \text{(A)}$$

where g_Y is the vector of underlying genotypes at the disease/trait locus, g_M, the vector of genotypes at the observed marker locus, and X is a vector of measured covariates. $P(g_Y, g_M)$ is the joint probability of the genotypes at the disease and marker loci and $P(Y \mid g_Y, X)$ is the penetrance function. Note that $P(M \mid g_M) = 1$ for a codominant marker. For individuals with no ancestors in the pedigree (founders and spouses), $P(g_Y, g_M)$ is

a function of the haplotype frequencies, *p(d-1), p(d-2), p(D-1),* and *p(D-2),* where *d* and *D* are the alleles at the disease locus and 1 and 2 are the marker alleles; the sum of these haplotype frequencies is constrained to be one. For individuals with ancestors in the pedigree, $P(g_Y, g_M)$ is a function of the Mendelian probabilities and the recombination fraction θ between the two loci. Because the dominantly inherited *CDKN2A* gene is here assumed to be the gene causing CMM in our family sample, there is complete identity between the disease and marker loci, which leads to $p(d\text{-}2) = p(D\text{-}1) = 0$ and $\theta = 0$. Thus, in this situation, when $g_M = g_Y$, the likelihood (A) becomes:

$$L\ (\text{family}) = \sum_{g_{Y\_unobs}} P(g_{Y\_unobs} | g_{Y\_obs}) \\ P(Y | g_{Y\_unobs}, g_{Y\_obs}, X) \quad \text{(B)}$$

where the summation is over all possible genotypes (g_{Y_unobs}) in subjects untyped for the marker conditional on the observed genotypes (g_{Y_obs}) in their family members. Thus, only one allele frequency needs to be estimated, which will be designated *CDKN2A−* for the allele not carrying the *CDKN2A* mutation, whereas *CDKN2A+* is the mutated allele. The likelihood was computed with the REGRESS computer program, which uses Eq. A for general situations. As stated above, Eq. A is equivalent to Eq. B when there is complete confounding between the putative disease gene and the marker, as is the situation here.

The penetrance function $P(Y \mid g_Y, X)$ over *n* individuals in the family is decomposed in a product of penetrance functions for each individual *i*: $P(Y \mid g_Y, X) = \Pi_{i=1}^{n}(P_i \mid g_{Yi}, Y_{ji}, X_i)$, where g_{Yi} is the *i*th person's genotype at the *CDKN2A* locus, Y_{ji} are the phenotypes of *j* antecedents of *i*, and X_i, the vector of covariates of *i*. Survival analysis concepts are introduced to take into account the variable age at diagnosis of CMM. The period of follow-up (taken here from 5 years of age because the risk of CMM is negligible before age 5), to age at diagnosis for affecteds, age at examination for unaffecteds (or affecteds with unknown age at diagnosis), or age at death for deceased subjects is partitioned into *K* mutually exclusive intervals. In each interval, we compute the hazard function, $\lambda(k)$, the probability of being affected in the *k*th interval given not being affected before. The penetrance function is then derived from λ(k). It is a density function at a given age of diagnosis included in the *k*th interval for affecteds, $f(k) = \lambda(k) \Pi_{h=1}^{k-1} [1 - \lambda(h)]$ and a cumulative function for unaffecteds with age at examination in the *k*th interval, $S(k) = \Pi_{h=1}^{k} [1 - \lambda(h)]$. For affected individuals with an unknown age at diagnosis and age at examination in the *k*th interval, the penetrance becomes $F(k) = 1 - S(k)$. The penetrance function is one for individuals with unknown disease status. The hazard function $\lambda(k)$ for the *i*th person in the *k*th interval is a logistic function $\lambda(k) = \exp[\theta_i(k)]/(1 + \exp[\theta_i(k)])$ where $\theta_i(k)$, the logit of the hazard function, is:

$$\theta_{i(k)} = \alpha_{gi} + \sum_{j=1}^{i-1} \Gamma_{ji} Y_{ji} + \beta_{gi} X_{i(k)} \quad \text{(C)}$$

where α_g is the genotype-specific baseline parameter, Γ_{ji} is a vector of regression coefficients on *j* antecedents' phenotypes of the *i*th person, and β_g is a vector of genotype-specific regression coefficients for covariates. We assumed a dominant mode of inheritance for the *CDKN2A* locus. The class D model specifies four types of family dependences of the *i*th person on his/her antecedents: spouse (Γ_{Si}), father (Γ_{Fi}), mother (Γ_{Mi}), and preceding siblings (Γ_{Ci}). The covariates included age and the three clinicoepidemiological covariates (TN, DN, and SI). Age was the only time-dependent covariate, whereas the effects of the others were assumed to be constant over time. Age was treated as a continuous covariate in the model, and different functions of age were considered, polynomial in age and natural logarithm of age. The logarithm function was found to fit the data better and was subsequently used in all analyses. In addition, no analyses showed evidence for an age by *CDKN2A* mutation interaction; therefore, all models presented include only one regression coefficient for the logarithm of age.

Hypothesis Testing. Parameter estimation and tests of hypotheses were carried out using maximum-likelihood methods as implemented in the program REGRESS (12, 13), which incorporates the regressive approach in the ILINK program of the LINKAGE package (16) and uses the GEMINI optimization routine (17). The joint likelihood of the CMM phenotypes and *CDKN2A* genotypes was maximized under different models, always including an autosomal dominant major gene effect at the *CDKN2A* locus. Incorporation of familial dependencies into the models showed results similar to the analyses that ignored familial dependencies but with less power to differentiate between models because of the increased number of parameters (data not shown). Therefore, our results are presented without residual familial dependencies. However, additional analyses were conducted by introducing the parental affection status as a covariate in the model to allow for aggregation of disease in the seven families without cosegregating *CDKN2A* mutations (as explained in the "Discussion"). Gene-covariate interactions were tested by comparison of submodels in which the βgs were set equal (no interaction) *versus* models in which two βgs (under a dominant model) were estimated (interaction). Effects and interactions of the three risk factors (TN, DN, and SI) with the major gene were tested separately and jointly.

Analysis Approach. Two types of analyses were performed to deal with missing/unknown covariates (18). The first approach (MI method) created two dummy variables for each covariate: a missing-value indicator set equal to 1 for missing/unknown and 0 for known; and a second variable set equal to 1 for exposed and 0 for unexposed or unknown. Thus, tests for gene-covariate interactions required comparison of four models: 1, β(MI) and β(E) for variables missing value indicator (MI) and exposure (E), both independent of genotype; 2, βg(MI) dependent on genotype but β(E) independent of genotype; 3, βg(E) dependent on genotype but β(MI) not dependent; and 4, βg(MI) and βg(E), both dependent on genotypes. If a test of model 1 *versus* model 2 was rejected, *i.e.,* a test of whether the effect of the missing covariate on disease risk was independent of genotype, the test of interaction for the exposure variable would compare model 2 *versus* model 4; otherwise, the test of interaction would compare model 1 *versus* model 3. The second approach to deal with missing covariates coded subjects with missing covariates as unknown for the disease status (the CC method). Therefore, for the CC method, tests for gene-covariate interaction required comparison of only β(E) and βg(E).

To search for gene-covariate interactions, we dichotomized the three previously identified risk factors (9) into referent/unexposed and risk/exposed categories. The baseline and at-risk categories for the covariates were defined as follows: TN, if age $<$16 years or age $>$50 years, $<$50 nevi *versus* $\geq$50 nevi, and if age 16–50 years, $<$100 nevi *versus* $\geq$100 nevi; DN, absent *versus* present; and SI, absent *versus* present. Table 1 presents a distribution of the covariates in each family.

Table 2 Combined segregation/linkage analysis of CMM, incorporating covariates TN, DN, and SI (MI method)

Model	Allele frequency[a] *CDKN2A*+	α(−)	α(+)	β(age)	TN			DN			SI			−2ln L
					TN β(−)	TN β(+)	UnkTN β*	DN β(−)	DN β(+)	UnkDN β*	SI β(−)	SI β(+)	UnkSI β*	
1. No sun-related covariates	0.06	−10.6	−8.4	1.5										1438.90
2. TN added to model	0.06	−11.0	−9.0	1.5	1.2	[=β(−)]	0.4							1413.23
2a. TN-*CDKN2A* interaction	0.06	−11.0	−9.0	1.5	1.4	1.2	0.4							1413.05
3. DN added to model	0.06	−12.3	−10.6	1.6				2.4	[=β(−)]	1.3				1386.55
3a. DN-*CDKN2A* interaction	0.06	−12.6	−10.2	1.6				2.8	1.8	1.0				1382.27
4. SI added to model	0.06	−10.8	−8.4	1.6							−0.2	[=β(−)]	−0.5	1435.47
5. TN and DN added	0.06	−12.4	−10.9	1.6	0.6	[=β(−)]	1.2	2.1	[=β(−)]	0.4				1373.03
5a. TN-*CDKN2A* inter; and DN	0.06	−12.3	−11.0	1.6	0.4	0.8	1.3	2.1	[=β(−)]	0.4				1372.48
5b. DN-*CDKN2A* inter; and TN	0.06	−12.6	−10.5	1.6	0.6	[=β(−)]	1.2	2.4	1.6	0.2				1370.78

[a] −, *CDKN2A* mutation absent; +, *CDKN2A* mutation present. α, genotype-specific baseline parameter [*e.g.*, α(−) is *CDKN2A* mutation absent baseline parameter]; β, regression coefficients of covariates [*e.g.*, TN β(+) is *CDKN2A* mutation present regression coefficient of TN]; β*, regression coefficient, not genotype specific.

Table 3 Combined segregation/linkage analysis of CMM, incorporating covariates TN, DN, and SI (CC method)

Model	Allele frequency *CDKN2A*+	α(−)	α(+)	β(age)	Total Nevi		Dysplastic Nevi		Solar Injury		−2ln L
					TN β(−)	TN β(+)	DN β(−)	DN β(+)	SI β(−)	SI β(+)	
1. No sun-related covariates	0.06	−10.9	−8.6	1.6							1134.88
2. TN added to model	0.06	−11.2	−9.1	1.6	1.1	[=β(−)]					1113.75
2a. TN-*CDKN2A* interaction	0.06	−11.3	−9.1	1.6	1.3	1.0					1113.51
3. DN added to model	0.06	−12.2	−10.7	1.6			2.4	[=β(−)]			1087.33
3a. DN-*CDKN2A* interaction	0.06	−13.0	−9.9	1.6			3.2	1.5			1083.36
4. SI added to model	0.06	−10.9	−8.5	1.6					−0.2	[=β(−)]	1134.25
5. TN and DN added	0.06	−12.2	−10.7	1.6	0.5	[=β(−)]	2.2	[=β(−)]			1082.75
5a. TN-*CDKN2A* inter; and DN	0.06	−12.2	−10.9	1.6	0.2	0.6	2.2	[=β(−)]			1082.25
5b. DN-*CDKN2A* inter; and TN	0.06	−13.0	−9.9	1.6	0.5	[=β(−)]	3.0	1.2			1078.51

[a] −, *CDKN2A* mutation absent; +, *CDKN2A* mutation present. α, genotype-specific baseline parameter [*e.g.*, α(−) is *CDKN2A* mutation absent baseline parameter]; β, regression coefficients of covariates [*e.g.*, TN β(+) is *CDKN2A* mutation present regression coefficient of TN]; β*, regression coefficient, not genotype specific.

Results

Table 2 shows the results of the combined segregation/linkage analyses for the MI method. Overall, there was a significant improvement in the fit of the model when TN (model 2 *versus* model 1: $\chi^2_2 = 25.7$, $P < 0.001$), DN (model 3 *versus* model 1: $\chi^2_2 = 52.3$, $P < 0.001$), or both nevi variables (model 5 *versus* model 1: $\chi^2_4 = 65.9$, $P < 0.001$) were included. The effect of either one of these variables was significant when the other one was included in the model. In contrast to DN and TN, inclusion of SI did not significantly improve the likelihood (model 4 *versus* model 1: $\chi^2_2 = 3.4$, $P = 0.19$). TN showed no evidence for gene-covariate interaction either alone (model 2a *versus* model 2: $\chi^2_1 = 0.2$) or with DN in the model (model 5a *versus* model 5: $\chi^2_1 = 0.6$). In contrast, there was significant evidence for interaction between DN and *CDKN2A* when DN was the only covariate in the model (model 3a *versus* model 3: $\chi^2_1 = 4.3$, $P = 0.04$). We also found that the regression coefficient for the unknown/known dummy variable was independent of genotype (data not shown). The DN-*CDKN2A* interaction was no longer significant when TN was added to the model (model 5b *versus* model 5: $\chi^2_1 = 2.2$). Interestingly, the estimate of the regression coefficient for DN (model 3a) was greater in subjects without *CDKN2A* mutations ($\beta = 2.8$) *versus* those with mutations ($\beta = 1.8$). The OR for DN in subjects without mutations was 16.4 (95% CI, 6.1–46.1) *versus* OR, 6.0 (95% CI, 2.3–15.5) among subjects with *CDKN2A* mutations.

The CC method (Table 3) showed results similar to the MI method with respect to addition of covariates to the model. Inclusion of DN, TN, or both significantly improved the likelihood (model 3 *versus* model 1: $\chi^2_1 = 47.6$, $P < 0.001$; model 2 *versus* model 1: $\chi^2_1 = 21.1$, $P < 0.001$; model 5 *versus* model 1: $\chi^2_2 = 52.1$, $P < 0.001$, respectively), whereas adding SI did not significantly improve the fit of the model ($\chi^2_1 = 0.6$, $P = 0.44$). TN revealed no evidence for interaction with *CDKN2A*. DN showed significant evidence for interaction with *CDKN2A* when it was the only covariate (model 3a *versus* model 3: $\chi^2_1 = 4.0$, $P = 0.046$) or when TN was also in the model (model 5b *versus* model 5: $\chi^2_1 = 4.2$, $P = 0.04$). This latter result contrasted with the MI analysis, where a significant DN-*CDKN2A* interaction was observed only when DN was the sole covariate in the model. For the CC method, the estimates of the regression coefficients for DN were similar to what was observed in the MI analysis; also, β(DN) was significantly greater in subjects without *CDKN2A* mutations (β − (DN) = 3.0) *versus* those with mutations (β + (DN) = 1.2). The OR for DN was 20.1 (95% CI, 4.8–92.8) in subjects without *CDKN2A* mutations compared with OR, 3.3 (95% CI, 1.1–10.0) among subjects with mutations.

Restriction of the analysis to the 13 families with cosegregating *CDKN2A* mutations produced very similar results. The same evidence for inclusion of covariates (addition of DN to the model, $P < 0.001$; TN: $P < 0.001$; DN and TN: $P < 0.001$; SI: $P = 0.40$) as well as DN-*CDKN2A* interaction [DN-*CDKN2A* interaction with DN as the only covariate: $\chi^2_1 = 3.9$, $P < 0.05$, β − (DN) = 10.0, β + (DN) = 1.5; DN-*CDKN2A* interaction with DN and TN as covariates: $\chi^2_1 = 4.2$, $P < 0.05$, β − (DN) = 10.2, β + (DN) = 1.2]. The estimate of the regression coefficient for DN in subjects with *CDKN2A* mutations (β+)

was similar to what was observed in the 20 families. However, in subjects without *CDKN2A* mutations ($\beta-$), this estimate, although higher than in the whole sample, should be regarded with caution because of the small numbers of CMM subjects without mutations.

Discussion

Combined segregation/linkage analysis of CMM in 20 American melanoma-prone families showed significant evidence for inclusion of the covariates DN and TN in the regressive models. A significant gene-covariate interaction between DN and *CDKN2A* was observed when DN was the only covariate in the model (MI and CC methods) or when both DN and TN were in the model (CC method only).

The *CDKN2A* gene has been implicated in melanoma pathogenesis from studies of melanoma-prone families, cultured and noncultured melanoma tumors, and functional studies of specific mutations. The pattern of *CDKN2A* mutations observed in cultured and noncultured melanoma tumors (19–22) indicates a possible role for UV radiation in the tumorigenic process. One hypothesis suggests that early loss of the p16 pathway may result in a net increase in melanocyte proliferation (23). This proliferation may then increase the probability that a melanocyte may accumulate additional damage during regular or UV-induced cell division, leading to development of a melanoma tumor. The finding of a significant association between an increased number of melanocytic nevi, the presence of DN, and the risk of melanoma are consistent with this proposed tumorigenic process.

Epidemiological studies of melanoma have demonstrated the importance of TN and DN as major risk factors for melanoma (reviewed in Ref. 7), yet few studies have been able to separate the effects of TN and DN. When TN and DN have been separated, the effect of DN appeared to be much greater than for TN (8). For this study, TN exposure was defined as an increased number of typical and atypical (dysplastic) nevi. DN exposure was defined by the presence of one or more unequivocal DN. Thus, there is overlap in the phenotypic definition of these two covariates. However, TN and DN were each found to influence separately the risk of melanoma, with DN showing an interaction with *CDKN2A* mutations.

Previous examinations of melanoma-prone families showed that DN did not appear to cosegregate with the *CDKN2A* mutations found in many families (14, 24, 25). The current study confirmed and quantified this relationship and showed that DN was a risk factor for melanoma, separate from *CDKN2A*; the risk for CMM associated with DN was greater in subjects without *CDKN2A* mutations compared with subjects with mutations. That is, among *CDKN2A* mutation carriers, the presence of DN multiplied the hazard function by 3 compared with mutation carriers without DN. In contrast, among noncarriers, the presence of DN multiplied the hazard function by 20 *versus* subjects without DN. In addition, the estimate of the hazard function among noncarriers with DN was similar to the estimate of the hazard function among *CDKN2A* mutation carriers without DN. Restriction of the analysis to the 13 families with cosegregating *CDKN2A* mutations showed a very similar pattern of results. In addition, because seven families did not have *CDKN2A* mutations, we performed additional analyses that allowed for aggregation of disease separate from *CDKN2A*. That is, we added covariates for the seven families based on whether a parent had CMM, was unaffected, or phenotype unknown. The "parental" covariates for the 13 *CDKN2A* families were coded as unknown. This analysis yielded results that were very similar to the CC analysis. We observed the same evidence for inclusion of DN and TN in the model as well as DN-*CDKN2A* interaction with comparable estimates of the regression coefficients (data not shown). Thus, although DN was a major risk factor for CMM in subjects with *CDKN2A* mutations, it appeared to be a stronger risk factor in subjects without *CDKN2A* mutations. This relative strength of effect for DN may change when the susceptibility gene(s) in the families without *CDKN2A* mutations are identified.

One potential problem in genetic epidemiological studies of families is that covariate data may be missing because not all family members are available or willing to participate in a research study. To deal with missing or unknown covariates, we conducted two analyses. The basic idea of the MI method is to obtain regression information from those family members for whom such information is available. In the CC method, family members with missing covariates were coded as unknown for the affection status. We observed similar results for the MI and CC method, although the CC method showed additional evidence for the DN-*CDKN2A* interaction. One difference between the two methods is that the MI method requires estimation of twice as many parameters for covariates, one covariate for known/unknown and one covariate for exposed/unexposed, as the CC method. This requirement, plus the observation that the DN-*CDKN2A* interaction was no longer significant after TN was added to the model in the MI method, raises a concern about adequate power for detecting gene-covariate interactions. It also raises concerns about possible biases in the methods. For example, the CC method may produce biased estimates of the parameters if the data are not missing completely at random. Missing completely at random is equivalent to assuming that, for each variable, the observed values effectively constitute a simple random sample of the values for all study subjects (26). For DN, similar percentages of missing values were observed in melanoma patients (27%) and those who did not have melanoma (28%). The MI method has also been shown to yield biased results in some situations and particularly if the covariate of interest is an important confounder of the study effect (26). Although the problems associated with missing data have been extensively evaluated for logistic regression (26), the problems have been rarely evaluated for family studies, where the issues may be more complicated and difficult to resolve. The use of more complex methods to allow for missing data, such as Gibbs sampling, have been proposed (27–29) but these methods have not yet been studied extensively. These issues will be further investigated.

A correction for ascertainment for these families was not applied to the likelihood calculation because this requires the mode of ascertainment to be known precisely. When it is unknown, as is the situation here, the MOD score strategy, which considers the likelihood of disease status and marker conditionally on the disease status of all family members, would be appropriate and lead to unbiased parameter estimates when there is no association between disease and marker (30, 31). In the situation of a disease-marker association, the MOD score should include linkage disequilibrium between disease and marker alleles not only in the numerator but also in the denominator. However, such conditional likelihood may remove most of the information, especially when complete linkage disequilibrium is specified, and may not be efficient. Although use of joint likelihood may affect genetic parameter estimates when the gene is unknown, this is less true when the gene is known in most subjects, as is the situation here. Although some error in inferring the gene carrier status in those subjects with unknown *CDKN2A* status adds noise, it is expected to decrease power for detecting gene-covariate interactions rather than causing a false-positive result.

The 20 families in the current study were ascertained because

they had at least two living first-degree relatives with melanoma. The average number of melanoma cases in the families was 5.5 (± 2.9). Although the families are not representative of all melanoma in the general population, they likely represent a rarer subset of melanoma-prone families with a strong genetic risk. These are the types of families that have been used to map melanoma and other disease genes (*e.g., BRCA1/2*). These types of families have yielded substantial information about the genetics of cancer and interactions with other genetic and environmental covariates. Thus, the results from these melanoma-prone families may be generalizable to that percentage of melanoma-prone families with strong genetic risks.

The etiology of melanoma is heterogeneous and complex. To date, two genes have been implicated in the inherited form(s) of melanoma. The major known gene *CDKN2A* confers substantial risk for melanoma. However, other genetic and environmental factors likely also contribute to disease expression. Interactions between a melanoma susceptibility gene and risk factors were suggested from multiple-case pedigrees from Utah where putative gene carriers, inferred from linkage analysis with 9p21 markers, had a higher number of nevi than noncarriers, and among gene carriers sun exposure was greater in subjects with than without melanoma (32). A significant interaction between a putative melanoma gene, detected by segregation analysis, and propensity to sunburn was reported in a large series of 295 melanoma French families ascertained through one melanoma proband (33). The current study of 20 American melanoma-prone families is the first to detect evidence for a significant interaction of a covariate with a known melanoma gene. The interaction between *CDKN2A* mutations and DN illustrate the complex etiology of melanoma. These results need to be confirmed in larger samples and in families from other geographic regions. In addition, the best approaches to deal with missing covariate data and the problems of adequate power to detect gene-environment interactions using the regressive models needs to be further assessed.

Acknowledgments

We are indebted to the participating families, whose generosity and cooperation have made this study possible. We acknowledge the Biomedical Supercomputing Center of the National Cancer Institute-Frederick Cancer Research and Development Center for allocation of computing time and staff support. We also acknowledge the contributions to this work that were made by Mary Fraser, Laura Fontaine, and Deborah Zametkin.

References

1. Greene, M. H., and Fraumeni, J. F., Jr. The Hereditary Variant of Malignant Melanoma. New York: Grune and Stratton, 1979.

2. Tucker, M. A. Individuals at high risk of melanoma. Pigment Cell Res., *9:* 95–109, 1988.

3. Serrano, M., Hannon, G. J., and Beach, D. A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/*CDK4*. Nature (Lond.), *366:* 704–707, 1993.

4. Wolfel, T., Hauer, M., Schneider, J., Serrano, M., Wolfel, C., Klehmann-Hieb, E., De Plaen, E., Hankeln, T., Meyer zum Buschenfelde, K-H., and Beach, D. A $p16^{INK4a}$-insensitive *CDK4* mutant targeted by cytolytic T lymphocytes in a human melanoma. Science (Washington DC), *269:* 1281–1284, 1995.

5. Zuo, L., Weger, J., Yang, Q., Goldstein, A. M., Tucker, M. A., Walker, G. J., Hayward, N., and Dracopoli, N. C. Germline mutations in the $p16^{INK4a}$ binding domain of *CDK4* in familial melanoma. Nat. Genet., *12:* 97–99, 1996.

6. Soufir, N., Avril, M-F., Chompret, A., Demenais, F., Bombled, J., Spatz, A., Stoppa-Lyonnet, D., Benard, J., and Bressac-de Paillerets, B. Prevalence of p16 and *CDK4* germline mutations in 48 melanoma-prone families in France. Hum. Mol. Genet., *7:* 209–216, 1998.

7. Armstrong, B. K., and English, D. R. Cutaneous malignant melanoma. *In:* D. Schottenfeld and J. F. Fraumeni, Jr. (eds.), Cancer Epidemiology and Prevention, Ed. 2, pp. 1282–1312. New York: Oxford University Press, 1996.

8. Tucker, M. A., Halpern, A., Holly, E. A., Hartge, P., Elder, D. E., Sagebiel, R. W., Guerry D., IV, and Clark, W. H. Clinically recognized dysplastic nevi: a central risk factor for cutaneous melanoma. J. Am. Med. Assoc., *277:* 1439–1444, 1997.

9. Goldstein, A. M., Falk, R. T., Fraser, M. C., Dracopoli, N. C., Sikorski, R. S., Clark, W. H., Jr., and Tucker, M. A. Sun-related risk factors in melanoma-prone families with *CDKN2A* mutations. J. Natl. Cancer Inst., *90:* 709–711, 1998.

10. Goldstein, A. M., Falk, R. T., Dracopoli, N. C., Sikorski, R. S., Clark, W. H., Jr., and Tucker, M. A. Clinical and epidemiologic factors in melanoma-prone families with p16 mutations. Genet. Epidemiol., *14:* 526, 1997.

11. Bonney, G. E. Regression logistic models for familial disease and other binary traits. Biometrics, *42:* 611–625, 1986.

12. Bonney, G. E., Lathrop, M., and Lalouel, J-M. Combined linkage and segregation analysis using regressive models. Am. J. Hum. Genet., *43:* 29–37, 1988.

13. Demenais, F., and Lathrop, M. REGRESS: a computer program including the regressive approach into the LINKAGE programs. Genet. Epidemiol., *11:* 291, 1994.

14. Hussussian, C. J., Struewing, J. P., Goldstein, A. M., Higgins, P. A. T., Ally, D. S., Sheahan, M. D., Clark, W. H., Jr., Tucker, M. A., and Dracopoli, N. C. Germline p16 mutations in familial melanoma. Nat. Genet., *8:* 15–21, 1994.

15. Abel, L., and Bonney, G. A time-dependent logistic hazard function for modeling variable age of onset in analysis of familial diseases. Genet. Epidemiol., *7:* 391–407, 1990.

16. Lathrop, G. M., and Lalouel, J. M. Easy calculation of LOD scores and genetic risks on small computers. Am. J. Hum. Genet., *36:* 460–465, 1984.

17. Lalouel, J. M., and Yee, S. POINTER: a computer program for complex segregation analysis with pointers. Tech. Rep., Population Genetics Laboratory, University of Hawaii, Honolulu, 1980.

18. Andrieu, N., and Demenais, F. Interactions between genetic and reproductive factors in breast cancer risk in a French family sample. Am. J. Hum. Genet., *61:* 678–690, 1997.

19. Healy, E., Sikkink, S., and Rees, J. L. Infrequent mutation of $p16^{INK4}$ in sporadic melanoma. J. Investig. Dermatol., *107:* 318–321, 1996.

20. Herbst, R. A., Gutzmer, R., Matiaske, F., Mommert, S., Kapp, A., Weiss, J., Arden, K. C., and Cavenee, W. K. Further evidence for ultraviolet light induction of CDKN2 ($p16^{INK4}$) mutations in sporadic melanoma *in vivo*. J. Investig. Dermatol., *108:* 950, 1997.

21. Luca, M., Xie, S., Gutman, M., Huang, S., and Bar-Eli, M. Abnormalities in the *CDKN2* (*$p16^{INK4}$/MTS-1*) gene in human melanoma cells: relevance to tumor growth and metastasis. Oncogene, *11:* 1399–1402, 1995.

22. Pollock, P. M., Yu, F., Qiu, L., Parsons, P. C., and Hayward, N. K. Evidence for U.V. induction of CDKN2 mutations in melanoma cell lines. Oncogene, *11:* 663–668, 1995.

23. Albino, A. P., and Fountain, J. W. Oncogenes and tumor suppressor genes in cutaneous malignant melanoma. Photochem. Photobiol., *63:* 412–418, 1996.

24. Gruis, N. A., Sandkuijl, L. A., van der Velden, P. A., Bergman, W., and Frants, R. R. CDKN2 explains part of the clinical phenotype in Dutch familial atypical multiple-mole melanoma (FAMMM) syndrome families. Melanoma Res., *5:* 169–177, 1995.

25. Wachsmuth, R. C., Harland, M., and Newton Bishop, J. A. Re: The atypical mole syndrome and predisposition to melanoma. N. Engl. J. Med., *339:* 348–349, 1998.

26. Greenland, S., and Finkle, W. D. A critical look at methods for handling missing covariates in epidemiologic regression analyses. Am. J. Epidemiol., *142:* 1255–1264, 1995.

27. Gauderman, W. J., and Faucett, C. L. Detection of gene-environment interactions in joint segregation and linkage analysis. Am. J. Hum. Genet., *61:* 1189–1199, 1997.

28. Gauderman, W. J., and Thomas, D. C. Censored survival models for genetic epidemiology: a Gibbs sampling approach. Genet. Epidemiol., *11:* 171–188, 1994.

29. Gauderman, W. J, Morrison, J. L., Carpenter, C. L., and Thomas, D. C. Analysis of gene-smoking interaction in lung cancer. Genet. Epidemiol., *14:* 199–214, 1997.

30. Risch, N. Segregation analysis incorporating linkage markers. I. Single-locus models with an application to type 1 diabetes. Am. J. Hum. Genet., *36:* 363–386, 1984.

31. Hodge, S. E., and Elston, R. C. Lods, Wrods, and Mods. The interpretation of Lod scores calculated under different models. Genet. Epidemiol., *11:* 329–342, 1994.

32. Cannon-Albright, L. A., Meyer, L. J., Goldgar, D. E., Lewis, C. M., McWhorter, W. P., Jost, M., Harrisson, D., Anderson, D. E., Zone, J. J., and Skolnick, M. Penetrance and expressivity of the chromosome 9p melanoma susceptibility locus (*MLM*). Cancer Res., *54:* 6041–6044, 1994.

33. Demenais, F., Chompret, A., Guilloud-Bataille, M., Feingold, N., Grange, F., and Avril, M. F. Interactions between genetic and epidemiological risk factors in cutaneous malignant melanoma in France. Am. J. Hum. Genet., *59 (*Suppl.*):* A28, 1996.